



That-clauses as complements of verbs or nouns

Document Version

Submitted manuscript

[Link to publication record in Manchester Research Explorer](#)

Citation for published version (APA):

Denison, D. (2018). *That-clauses as complements of verbs or nouns*. In E. Seoane, C. Acuña-Fariña, & I. Palacios-Martínez (Eds.), *Subordination in English: Synchronic and diachronic perspectives* (pp. 61-84). (Topics in English Linguistics). de Gruyter, Walter GmbH & Co.

Published in:

Subordination in English

Citing this paper

Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights

Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy

If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [<http://man.ac.uk/04Y6Bo>] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.



Not final version. Please do not quote without permission.

To appear in *Subordination: Synchronic and diachronic perspectives* (Topics in English Linguistics). Berlin: Mouton.

That-clauses as complements of verbs or nouns

David Denison
University of Manchester

Abstract

Many types of verb in English permit a *that*-clause complement, including factual and suasive verbs like *acknowledge*, *point out*, *recommend*, *suggest*, etc.:

- (1) I assume she acknowledges that it was a blunder (2015, COCA)

I refer to this construction as 'V + *that*'.

Other transitive verbs like *advance*, *contest*, *contradict*, *endorse*, *highlight*, *moot*, *pose*, *put forward*, *propound*, *uncover* are not supposed to take *that*-clauses, and by and large in the pre-2000 corpora they don't. Instead such verbs take a nominal object, one form of which can be a 'shell noun' (abstracts like *fact*, *claim*, *argument*, *situation*) with a *that*-clause as complement of the noun, a construction referred here to as 'shell + *that*'. Both a simple NP and an NP containing a *that*-clause are illustrated in (2):

- (2) And I think Lord Scarman's right to highlight the era of human rights and the fact that the police are the key institution in our society (DCPSE)

In recent English, simple *that*-clause complements are spreading to such verbs in the active, (3)

- (3) a. In the first issue of *Salvage*, Neil Davidson mooted that neoliberalism may be undermining the basis for capital accumulation itself. (2015, China Miéville, 'On Social Sadism', *Salvage* <http://salvage.zone/in-print/on-social-sadism/> 17 Dec 2015)
b. The renowned Millennium Ecosystem Assessment [...] highlights that coasts are the most highly degraded ecosystems on our planet. (2005, COHA)

In this paper I explore data from COHA bearing on the development from shell + *that* to V + *that*, discussing the difficulties of getting reliable evidence. The topic also raises a tricky problem for corpus linguists, namely whether it is ever legitimate to talk of native speakers not being in full control of some registers of their own language.

That-clauses as complements of verbs or nouns¹

David Denison
University of Manchester

1 What

I begin by explaining *what* usage is to be discussed (Section 1), move on to *how* it might have come about (Section 2), detail my corpus investigation into *when* it appeared (Section 3), then raise questions about *whose grammar* is involved and what that might mean (Section 4), before a brief consideration of *what else* needs doing (Section 5).

Many verbs of communication, factual and suasive verbs can include a finite-clause proposition in their complement, either directly as a *that*-clause, (1), or indirectly with a *that*-clause dependent on an object noun, as in (2):²

- (1) claim that X, say that X, suggest that X
- (2) advance the claim that X, formulate the suggestion that X, welcome the fact that X

Such object nouns have been defined in various ways and given a number of different labels, most recently ‘signalling nouns’ (Flowerdew & Forest 2015), though I will stick to the more widely used ‘shell nouns’ (Hunston & Francis 2000: 185-6, Schmid 2000).³ My concern here is with the spread to new verbs of pattern (1) complementation (henceforth ‘V + *that*’) at the expense of (2) (‘shell + *that*’). I present some evidence of recent change and provide preliminary answers to where, when, how and why. For this I conduct a systematic investigation in the Corpus of Historical American English (COHA, Davies 2010-).

According to Biber et al. (1999: 668), ‘... communication verbs controlling *that*-clauses (apart from *say*) are most frequent in academic prose’, and indeed my first exposure to the usage which is the topic of this paper was in essays by British students. Many of the writers were relatively unskilled, insecure about written expression, and they may have been using a word-processor thesaurus for ‘elegant variation’ or to avoid the risk of plagiarism. Here are some genuine examples which contain a verb used directly with a *that*-clause complement that to my ears needed an NP complement. (The names of cited scholars are replaced by X to avoid distraction.)

¹ Early versions were presented at ICLCE3 (London, 2009) and ISLE2 (Boston MA, 2011), more briefly as part of lectures at Aoyama Gakuin University (Tokyo, 2016) and University of Hiroshima (2016), and at LMEC6 (Uppsala, 2017). I thank all those audiences for a number of insightful comments.

I am deeply grateful to Jiří Zámečník, a member of research training group “Frequency effects in language” at Freiburg, who generously provided the basis of two Python scripts for the offline COHA files. He helped to debug and improve the scripts, optimising them to run 40 times faster than at first and to handle errors encountered in the files, and he pointed out the unreliability of a conclusion based on low frequencies.

² There are, of course, introducers of finite clause complements other than *that*, notably *whether* and *how*.

³ Hunston & Francis (2000: 186) characterise shell nouns semantically as follows:

The idea that they are a distinct and definable group – arguably a class of their own – is reinforced by the discovery that all these nouns have features of meaning in common. They fall into two major groups. The first group consists of nouns which refer to something that is written or spoken; the *that*-clause lexicalises what it is that is written or spoken. The second group consists of nouns which refer to beliefs, ideas, wishes, and thought processes. Again, the *that*-clause lexicalises what it is that is thought or believed.

- (3) X et al. (1985) advance that the *try* in *try and V* is ... (2009)
- (4) X (1970) notes how this order of events is valid, citing that ... (2015)
- (5) Along with X who contested that people who reached for [...] (2012)
- (6) [...] with X criticising that the French influence was sporadic [...] (2010)
- (7) X defines that “grammar may be regarded either from a theoretical or practical point of view. [...]” (2011)
- (8) Within this discussion he deliberates that “‘Correctness’ as it is popularly conceived, is one of these terms that seems to refer to something desirable, [...]”. (2011)
- (9) The OED entry depicts that the word was originally borrowed from French ... (2014)
- (10) X expressed that Middle English was indeed a creole [...] (2010)
- (11) John Cheke who was a professor of Greek at the University of Cambridge instigated that our language should be written cleane & pure’. (2012) [punctuation as shown; intended meaning of verb unclear – DD]
- (12) This portrays that the preacher deserves pity. (2011)
- (13) X poses that for males the same interruptions may act as a bid for the floor (2013)
- (14) [...] which can be reinforced by X et al, who utters that, “In other locations [...]” (2010)

Since I gave presentations on this phenomenon in Denison (2009, 2011), I have collected further examples, often from journalism, to expand the range of verbs involved. Nevertheless that collection is unsystematic and limited to a couple of examples of each verb. As for the supposed standard Present-Day English (PDE) against which I was comparing this usage, I tried to gauge the scholarly consensus by combining relevant lists of verbs in Biber et al. (1999: 362-3, 660-70), Huddleston & Pullum (2002: 958-9, 1017-22), Quirk et al. (1985: 1176-7, 1180-83), Collins COBUILD (1987).⁴ I also conducted searches in the *OED* and a number of corpora for a few specific verbs.

Verb subcategorisation in English is partly systematic but in part apparently arbitrary. Huddleston & Pullum, for example, devote three chapters to different patterns of complementation (2002: Chapters 4, 11, 14), offering a combination of explanatory principles and mere lists. The other major grammars (Quirk et al. 1985, Biber et al. 1999) have comparable expositions. Dictionaries organise similar material by verb rather than by complementation pattern. Possible complementation patterns are a particular focus of learner dictionaries.

My tabulation of information in the handbooks is as shown in Table 1. The first group of verbs are reported as allowing either shell + *that* or V + *that*, while the second group have limited occurrence in shell + *that* that is not synonymous with use in V + *that*. The third group are reported as allowing shell + *that* but not V + *that* (though *dispute* and *question* can occur with a *whether*-clause). I have marked in bold those that I have attested with a *that*-clause – *contra* the handbooks. The relevant behaviour of the fourth group is not specifically mentioned in the handbooks (apart from *inform* occurring with an indirect object and *that*-clause) and therefore is in particular need of corpus study. These are verbs which I have attested in V + *that* but which in 2011 I felt would not be possible in my idiolect. The unreliability of introspection is such that I would no longer reject with any certainty *deem*, *posit*, *reiterate* or *take into account*.

⁴ Note that REPORT-CL in COBUILD covers more than just *that*-clauses.

verb	shell + that	V + that
<i>accept, acknowledge, add, affirm, allege, allow, announce, assert, assume, believe, concede, confirm, consider, convey, demonstrate, deny, disclose, discover, doubt, emphasise, establish, explain, forget, guarantee, hold, imagine, infer, maintain, mention, observe, ?pronounce, propose, prove, recognise, regret, repeat, report, see, state, stress, submit, suggest, suspect, understand</i>	✓	✓
<i>claim, deduce, determine, find, indicate, judge, point out, predict, presume, show, stipulate, suppose</i>	? (≠)	✓ (≠)
<i>advance, articulate, back up, challenge, communicate, contradict, convey, define, discuss, dispute, encourage, endorse, enlarge upon, espouse, express, oppose, promote, put across, put forward, question, rule, support, sustain, underline, underscore, utter</i>	✓	×
<i>advocate, analyse, bring to the surface, cite, clarify, contest, criticise, deem, deliberate, depict, describe, display, exemplify, explicate, highlight, identify, illustrate, inform, instigate, interpret, moot, portray, pose, posit, propound, publicise, quote, reflect, refute, reinforce, reiterate, respect, rule out, solidify, stand, summarise, take into account, uncover, update, view, welcome, yield</i>	?	?

Table 1: Complementation by shell + *that* and/or by *that*-clause alone

2 How?

The verbs that can occur in patterns (1) and (2) are by no means mutually exclusive sets. Many verbs (listed in the first row of Table 1) take part in both patterns with very similar meanings either way:

- (15)a. accept that X, disclose that X, prove that X, understand that X
- b. accept the claim that X, disclose the fact that X, prove the case that X, understand the observation that X

This point is clearly crucial to any explanation of the spread of the V + *that* construction. The simplest mechanism responsible would be analogy: the dual subcategorisation of verbs like those in (15) is extended to verbs like those in (2). With sufficiently detailed information on numbers and timing, an analogical account could be reframed more precisely in Construction Grammar terms.

If analogy is a mechanism at work here, it raises the question of timing. Why should an analogy which has existed for centuries prove so productive only now? Has the spread merely become more noticeable because we now have access to more unedited material from less experienced writers, or is it that less experienced writers are taking advantage of online thesauri which offer a variety of verb synonyms but are silent on their complementation patterns?⁵ Then again, is the development in question actually as recent as it seems? Perhaps that impression was mistaken, and there is a longer-term tendency for verbs to add the possibility of *that*-clause complementation.⁶

⁵ A resource like the *Oxford Advanced Learner's Dictionary* would give usage guidance on a verb, while the Language Bank on its website could suggest synonyms, but I don't know how long such services have been in existence online. In any case I doubt that many British native speakers would think of using a resource explicitly aimed at language learners; see Section 4.

⁶ This possibility was underscored by two examples from the Sherlock Holmes stories of Conan Doyle provided by Professor Minoji Akimoto after my lecture in Tokyo (Nov. 2016):

(i) I even distinguished that one of them was marked with the name of 'Hyams', who was Oldacre's tailor. (1903, *The Adventure of the Norwood Builder*)

Zwicky (2005) has two salutary warnings here:

[...] the Recency Illusion, the belief that things YOU have noticed only recently are in fact recent. This is a selective attention effect. Your impressions are simply not to be trusted; you have to check the facts. Again and again -- retro *not*, double *is*, speaker-oriented *hopefully*, split infinitives, etc. -- the phenomena turn out to have been around, with some frequency, for very much longer than you think. It's not just Kids These Days.

Professional linguists can be as subject to the Recency Illusion as anyone else. [...] Another selective attention effect, which tends to accompany the Recency Illusion, is the Frequency Illusion: once you've noticed a phenomenon, you think it happens a whole lot, even "all the time". Your estimates of frequency are likely to be skewed by your noticing nearly every occurrence that comes past you. People who are reflective about language -- professional linguists, people who set themselves up as authorities on language, and ordinary people who are simply interested in language -- are especially prone to the Frequency Illusion.

Note in addition that the simple analogical explanation would not cover *deliberate that X* in (8), since *deliberate* is an intransitive verb which does not occur in shell + *that*, while the sense of *instigate* in (11) is so obscure that the example should probably be rejected outright as an error. However, the whole question of grammaticality vs. error is central to this investigation and cannot be taken for granted; I take this point up in Section 4.

Returning to the question of how new V + *that* usages might have arisen, the simple analogy mentioned above is not the only possible mechanism for some verbs. Newmeyer (2003: 160 n.12) is one scholar who has pointed out that

A number of verbs disallow sentential complementation in surface direct object position, but do allow them as subject of passives:

- (i) a. *The grammar expresses [_{CP} that the rule is obligatory].
b. [_{CP} That the rule is obligatory] is expressed by the grammar.
- (ii) a. *This formulation of the rule reflects [_{CP} that all NPs behave uniformly]
b. [_{CP} That all NPs behave uniformly] is reflected by this formulation of the rule.
- (iii) a. *This theory captures [_{CP} that languages are learnable].
b. [_{CP} That languages are learnable] is captured by this theory.

Newmeyer's examples have a *that*-clause as passive subject, a structure which is rather high-flown. Extraposed passive examples with a dummy *it*-subject are less rare, (16), and passive participles also occur in the complement of other verbs, as in (17); see also (22)-(24), (29) below.

(16) but it must be faced that I have doused all hope of continued revelry (1951, COHA)

(17) Don't you want it known that you're all right? (1921, COHA)

(ii) And in this way he managed that your good man should have no want of drink [...] (1892, *The Adventure of the Copper Beeches*)

In the light of these facts, a conceivable route to some new cases of *V + that* would be for a subcategorisation already licensed for a passive matrix verb to be extended to the same verb in the active. Although the investigations described in Section 3 threw up a modest number of passive examples, they were not numerous enough to test relative frequency statistically, so their possible relevance to the entry and spread of new *V + that* patterns remains unknown. Intuitively, though, what are often rather elaborate, formal-sounding passives are rather foreign to the style at least of the student writers we have mentioned, making it unlikely that the passive would have been an important conduit for their usage.

For some of the verbs where *V + that* is unexpected, Erik Smitterberg made a valuable comparison between a *that*-clause and a finite clause introduced by *how*, speculating whether the existence of *V + how* could somehow have helped to license *V + that*. I give parallel examples in (18) and (19):

- (18)a. I love how they're always using the phone. (1970, COHA)
- b. but he couldn't stand how she refused to take her own life seriously (1989, COHA)
- (19)a. I love that she kept some of the funny wallpaper in her house (2004, COHA)
- b. and then you couldn't stand that I was becoming successful (2002, COHA)

While most *how* subclauses have a manner or degree reading in line with the usual meanings of *how*, there are some – such as those I have cited in (18) – where the meaning of *how* approaches the purely factitive meaning of *that*. A possible route to the spread of *V + that* would then be as follows: manner *how* gets bleached to a neutral subordinator and is then replaced by the more general subordinator *that*.

We might go further and compare *the way (that)*, which is semantically similar to *how* in allowing both a manner reading and a bleached, factitive meaning, as in

- (20)and I resented the way that accidents had snatched their dignity from them (2006, COHA)

Then we would have a route from shell + *that*, as in (20), to *V + how*, as in (18), to *V + that*. I have not explored scenarios involving *how* in any detail, but likely problems include relative dates and low numbers: I suspect that some supposed precursors are no older than the *V + that* patterns to be explained, and it looks as if types (verbs which can subcategorise for factitive *how*) and tokens are both infrequent. For this paper I concentrate on the direct analogical route from shell + *that* to *V + that*.

3 When?

This section concerns the timing and frequency of the linguistic development, which inevitably involves a detailed description of the methodology employed.

3.1 Problems with an online search

To see how reliable the handbooks are in their characterisation of what is and isn't possible, we need a baseline survey of PDE corpora. And we need to survey historical corpora to document the advent and rise of *that*-clause complementation with verbs that had not previously permitted it. String searches would be too inefficient: many of the verbs in question are identical in form to nouns, and the word *that* is too multifunctional. I therefore start with a large tagged historical

corpus, COHA. As far as I am aware, this is currently the only corpus with historical coverage up to the present which is large enough to provide a sufficiently wide range of data.

A search for **[display].[v*] that.[cst]** in COHA specifies the verb *display* in any of its forms and the conjunction *that* (which includes the relativiser function). Of 41 hits, 18 have the noun *display* wrongly tagged as a verb,⁷ and another 18 have determiner *that*. Now the accuracy of tagging in COHA leaves something to be desired, and here we are looking for a construction which may be innovative for the verb *display* and absent from the tagger's training corpus, but an accuracy of 5/41 or 12% is too low to be helpful. Furthermore, of those five *that*-clauses following the verb *display*, one is not in construction with that verb, and three are relative clauses, so in the end only one is at all relevant:

- (21)The man made the animal open its mouth as wide as possible to appear fearsome, but all this accomplished was to display that the poor creature's teeth have been yanked out.
(1990, COHA)

A similar search for **[advance].[v*] that.[cst]** produces 137 hits, of which 56 have the probable adjectival participle *advanced*, 13 have the noun *advance*, 2 have determiner *that*. Of the 66/137 correctly tagged (48%), 58 are irrelevant, leaving eight cases for consideration. Of these, three involve a passive with *that*-clause as extraposed subject.

- (22)We hav[e] nowhere seen it advanced that even a minority of the Legislature of 1839 made any movement of Repudiation which might at least have served as a notice of intention.
(1842, COHA)
(23)Since Hannah's death I have heard it openly advanced that she was the guilty party in the crime (1878, COHA)
(24)It has been advanced that great queens owed their power to the association and advice of the noble and high-minded men who surrounded them (1911, COHA)

And five examples, all early, do seem to have the verb *advance* in the active with a *that*-clause complement:

- (25)Buffon advanced that the brain of the orang-outang " does not differ from that of man
(1835, COHA) [writer of this and two other examples is called Franz Josef Gall]
(26)If some of them /q/ have advanced that every thing proceeded from corruption, others, more honest and sincere, have refuted them, even in the earliest times. (1836, COHA)
(27)Some have advanced that the finest roses from seed are always the longest in flowering
(1844, COHA)

Finally here, in the hope of higher precision I ran a similar search for a verb, *define*, which does not have a noun homonym. There are 42 hits for **[define].[v*] that.[cst]**, of which six have the probable or possible adjectival participle *defined*, three have pronominal *that*, nine have determiner *that*, and one is removed by the software as a probable duplicate. Of the 23/42 remaining (55%), two have relativiser *that* and 15 are irrelevant, leaving four early examples of *define* + *that*-clause, (28), one with passive *define* and extraposed *that*-clause subject, (29), and one fairly recent occurrence, (30):

⁷ However, a search for **[display].[n*] that** (the noun *display* followed by any *that*) produces 44 hits, while **[advance].[n*] that** gives 326, and every single example indeed has a noun and not a mistagged verb. This will be important in §3.3; see note 9.

- (28)The Pope defined that it was a revealed truth that Mary had been conceived without incur ring original sin (1861, COHA + 3 other examples in the same text)
 (29)It is defined that Almighty God is auctor utriusque Testamenti. (1882, COHA)
 (30)How to define that it has a kind of off-reddish tint, neither quite one color nor another, stumps me until [...] (1993, COHA)

From these brief experiments with tagging we see that precision in COHA searches for this kind of material can be poor – especially when the construction might be a recent innovation. To gauge recall, we would have to repeat the test for the lexemes *display/advance/define* with tags other than verb or with *that* tagged as something other than conjunction – In effect, we would have to discard the tagging.

Given the difficulty of conducting an efficient search in online COHA or a sister corpus like GBooks, I have considered recourse to parsed corpora. Possibilities include (i) the Penn Parsed Corpus of Modern British English, 2nd edition, release 1 (PPCMBE2), for the period up to 1914, and (ii) the Diachronic Parsed Corpus of Spoken English (DCPSE), to compare the 1960s and 1990s, in both of which the part of speech (POS) tags are manually verified. There is also (iii) the dependency-parsed version of ARCHER 3.2 at Zurich, for the period 1600-1999. However, because of their pre-PDE dates and modest sizes, none of these parsed corpora would permit the testing of the suspicion that online thesauri play a role in the spread of V + *that*, so I am thrown back on COHA for its size and its perfect chronological range, but necessarily the offline version.

3.2 Identifying verbs in V + *that*

We need a systematic, corpus-based means of identifying verbs that appear in shell + *that* and which might also allow V + *that*. Rather than collecting every possible *that*-clause complementing a verb, which would have been an overwhelming task, I started instead from the presumed source of the innovation, pattern (2), shell + *that*. I also limited my research to the twentieth century plus the first decade of the twenty-first, thus the last 11 decades of COHA.

In Flowerdew & Forest (2015) there are convenient lists of the shell nouns found in an academic corpus of PDE. I noted the 50 commonest.⁸ The procedure I eventually settled on was to search the largest and most recent decade of COHA, 2000-2009, for shell + *that* involving those shell nouns. The search, programmed in Python, strips out most punctuation and looks for suitable strings, with additional words of context to left and right. In effect – I translate the actual Python formulation into more familiar terms – the crucial portion looks for the following sequence in the corpus:

(31)verb + article + shell-noun + *that* ...

The program only uses POS tags to pick out ‘verb’⁹ and ‘article’ in (31) and to ensure that the verb is not *be*; the slot ‘shell-noun’ draws in a list of items, and the unreliable POS tagging of *that* is ignored.

⁸ The 50 commonest shell nouns listed in Flowerdew & Forest (2015: 86-7, Table 8.2) in descending order of frequency are *case, way, problem, result, thing, theory, model, idea, example, effect, point, reason, equation, question, process, issue, method, right, fact, principle, condition, approach, argument, procedure, change, difference, strategy, situation, solution, time, policy, analysis, view, decision, evidence, possibility, role, function, area, expectation, concept, ability, factor, conclusion, need, consequence, section, stage, failure, response*. I added the alternative spelling *rôle*.

⁹ From my experiment with *display* (note 7 above), it seemed unlikely that there would be many nouns mistagged as verbs. The reverse error is common.

On the first run I used the 20 most frequent shell nouns. The resulting hits were relatively quick to filter, leaving 604 (64% precision) that did exemplify shell + *that*. I sorted these by matrix verb to find the most common verbs introducing shell + *that*. Figure 1 shows those verbs occurring five times or more.

At first I investigated those verbs with six or more occurrences (left of the vertical line in Figure 1).¹⁰ When I came to see how often they occurred in V + *that* over the period 1900-2009, the results were disappointing: 1745 good examples (precision 73%), with individual totals very low for all but four verbs; indeed three out of 20 had none at all, while six more occurred ten times or less in V + *that*. Nor was there much in the way of chronological patterning. Selecting verbs on the basis of probable frequency in shell + *that* had not been particularly enlightening.

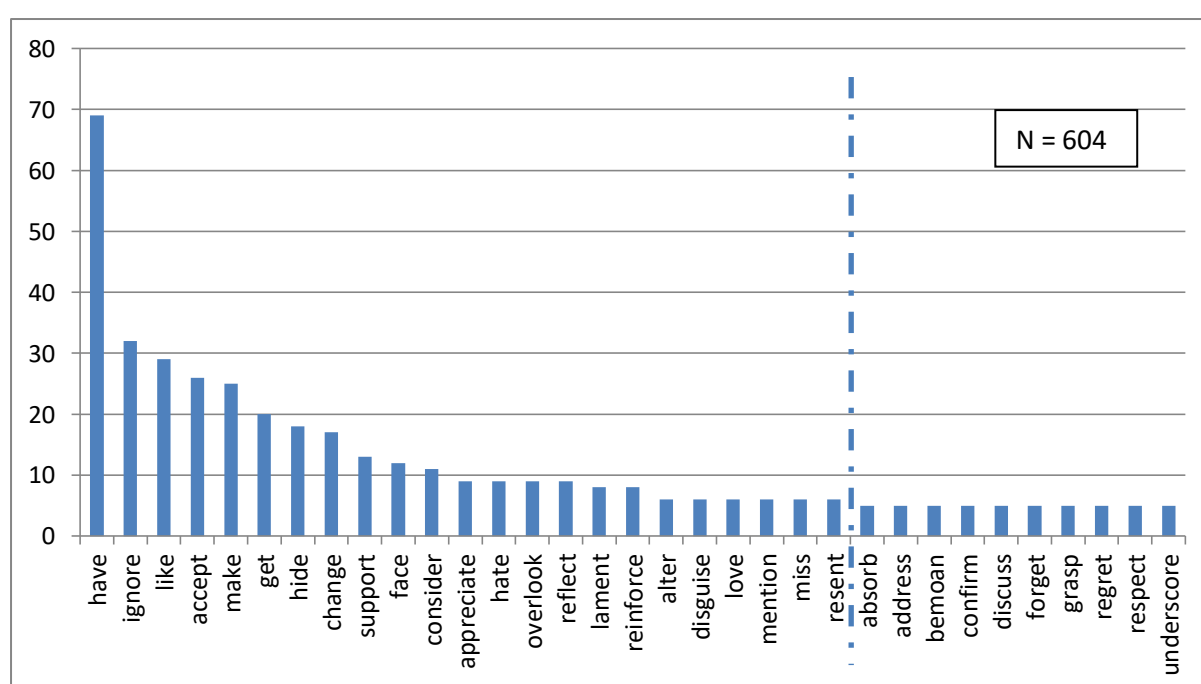


Figure 1: Most common verbs with shell + *that* 2000-2009

To make the list of suitable verbs more comprehensive, I added all remaining verbs with between 1 and 5 genuine examples of shell + *that* from the initial shell noun search: 136 verbs,¹¹ of which 51 would actually turn out to appear in V + *that* 1900-2009. I also ran the shell noun search again with nouns 21-50 from the list in note 8, once again only for the 2000s decade, but this time quickly sifting the hits for just one valid example of shell + *that* with any matrix verb not yet included in the research. This added a further 36 verbs, of which 21 later appeared in V + *that*, giving a grand total of 89 verbs which had occurred in COHA in shell + *that* at least once in the decade 2000-2009. Finally, from my own collection of attested examples with 'unexpected' (to me) V + *that* rather than

¹⁰ The 20 verbs are *accept, alter, appreciate, change, consider, disguise, face, hate, hide, ignore, lament, like, love, mention, miss, overlook, reflect, reinforce, resent, support*. The common *get, have* (generally occurring with shell *idea*) and *make* (with *case* or *point*) were excluded as not offering any synonymous V + *that* possibility.

¹¹ The list actually fed to the program includes a few extra verb lemmas to allow for spelling variants like *criticise/criticize, savor/savour*, as COHA is inconsistent about lemmatising such pairs together.

the normal shell + *that* , I added 15 more verbs, 9 of which appeared in the V + *that* searches of COHA.

3.3 Counting V + *that*

The search procedure for V + *that* was broadly similar to that for shell + *that*, except that this time it proved necessary to increase the initial precision (thus lowering the recall) by restricting what follows *that* to some of the more probable complement *that*-clause shapes:¹²

- (32)a. matrix-verb + *that* + pronoun + V ...
- b. matrix-verb + *that* + existential-*there* + V ...
- c. matrix-verb + *that* + N + V ...
- d. matrix-verb + *that* + N + N + V ...
- e. matrix-verb + *that* + article/possessive + N + V ...
- f. matrix-verb + *that* + determiner + N + V ...

The program is fed a list of items for the 'matrix-verb' slot; the word *that* is just a string; and the remaining items use the POS-tags of COHA. The rough-and-ready characterisation in (32) of likely *that*-clause openings successfully cuts out many examples which are not V + *that*, though of course at the cost of some good examples, e.g. those where the subject NP of the *that*-clause includes a post-modifier, or where an adverbial intervenes after *that*.

Compare an online search like **[respect].[v*] that**, where the verb is tagged but *that* is not. From a few experiments with different verbs, it looks as if such an online search typically retrieves about 2.5 times more valid V + *that* examples than the Python program described just now. With an online string search like **respect|respects|respected|respected|respecting that**, the haul of V + *that* might possibly be slightly greater still, but that would have to be set against the fact that with this verb, every single form has a non-verb homonym. Use of the verb tags is certainly worth it. Both offline and online searches make use of tagging of the matrix verb, so both would miss an example if the verb had been mistagged as a noun. In practice, neither the COHA server nor I have the patience for what would have been over 170 online searches, and once there are many thousands of hits to deal with, precision becomes crucial. For verbs that are frequent in V + *that*, the Python search summarised in (32) shows almost 100% precision and was the only practical choice. My assumption is that the sample of V + *that* obtained will nevertheless give a fair picture of chronological trends.

So far – and indeed throughout this paper – I have only allowed for cases where the *that*-clause follows on immediately from the shell noun in shell + *that* or the matrix verb in V + *that*. With V + *that*, therefore, my searches will not pick up examples like

- (33)a. That they are fair I can testify (1900, COHA)
- b. That the argument is sound can be seen by exhibiting its structure (1973, COHA)
- (34)a. You mentioned earlier that he'd summoned you. (1996, COHA)
- b. but he remembered with a slight wince that it had cost \$1,800 (1953, COHA)

¹² Jiří Zámečník's programming makes ingenious use of CLAWS tags to avoid the processing in Python of lengthy regex formulas. A similar restriction was tried out in the shell noun search but was found to lower recall more than it increased precision.

I will simply ignore pre-placement of *that*-clauses as in (33), on the assumption that any small increase in recall – for they are not common – would be vastly outweighed by a loss of precision; that is, if the search could be programmed at all.

What about *that*-clauses that follow their head but with something interrupting, as in (34)? I tested this for both early stages of my procedure. I ran a string search in online COHA for the noun *case* followed within nine words by *that* and took a sample of 200 from the 13015 hits. The sample contained 11 examples of shell + *that*, in ten of which *that* was adjacent to *case*, as against just one clear example where it was non-adjacent.¹³ Similarly, I searched for the verb lemma *accept* followed within up to nine words by *that*, this time in the period 1900-2009, and took a random sample of 200 from the 5402 hits: 25 out of 200 exemplified V + *that*, and in 23 of those 25, V and *that* were adjacent. It does seem reasonable, then, to limit searches to adjacent *that*.

One danger of confining searches to *that*-complements that immediately follow their head is that syntactic change is often thought to begin ‘sneakily’, in contexts where a change is least noticeable (De Smet 2012: 605). If such is the case here, there is a risk of missing some crucial early examples. But statistically and for reasons of sanity, the simplification of search strategy is justified.

All searches for V + *that* reported here were made with the same Python script as sketched in (32). The program takes a list of verbs and works with one decade of COHA data at a time. Most of the interesting chronological changes involve very low frequencies, but numbers of hits can be very large: one run produced 5528 for the 1900s alone, with many of the verbs showing hundreds of examples each in that decade: *see* ×1240, *realize/-ise* ×427, *declare* ×422, *admit* ×288, and so on. These numbers would overwhelm and mask any of the innovatory uses. What it confirms is that V + *that* is already deeply entrenched by 1900.

I decided to split the list of verbs at an arbitrary threshold, but one strongly suggested by the data. Any verb which occurred in V + *that* more than 20 times in both the 1900s and 1910s was treated as entrenched, and numbers of hits for all of them were combined – between 4700 and 6800 per decade – in order to provide a baseline measure of the frequency of V + *that* over the period. I could only sample the hits for these high-frequency verbs, but after scanning hundreds I had only spotted one non-example, so it seemed safe to count without filtering.

For all but a couple of the remaining verbs in the list, there were few or no hits for V + *that* in the 1900s or 1910s, with most hits that did occur having to be rejected. This is not unexpected for a usage which is recent in PDE and of dubious grammaticality. For these verbs, therefore, precision is unreliable, and all hits over the whole period were filtered for genuine cases of V + *that* and counted separately.¹⁴

¹³ I discounted a second, doubtful non-adjacent example, which probably involves complementiser *so that*:
(i) I say then, that, even though the case could be so that the whole system of Catholicism was recognized and professed, without the direct presence of the Church, still this would not at once make such a University a Catholic Institution [...] (1852, COHA)

¹⁴ Naturally, decisions have to be made about what constitutes a valid example of V + *that*. I discounted examples where a participle has developed into a conjunction (‘Given that ...’, ‘Providing that ...’), and such cases as ‘but it afterward developed that he was never near the place’, ‘It registers that Amina is no longer in the car’, where dummy *it* is subject of an active verb.

3.4 History of V + *that* 1900-2009

I present the results for different groups of verbs. The 24 'entrenched' verbs (*acknowledge, add, admit, assert, consider, declare, demand, explain, fear, find, forget, indicate, learn, mention, propose, protest, prove, realize, recognize, reflect, remember, see, suggest, understand*) are shown in Figure 2: The total frequency per million words falls somewhat after the 1920s but remains relatively high by comparison with most of the other verbs in the construction. There was no individual verification of examples.

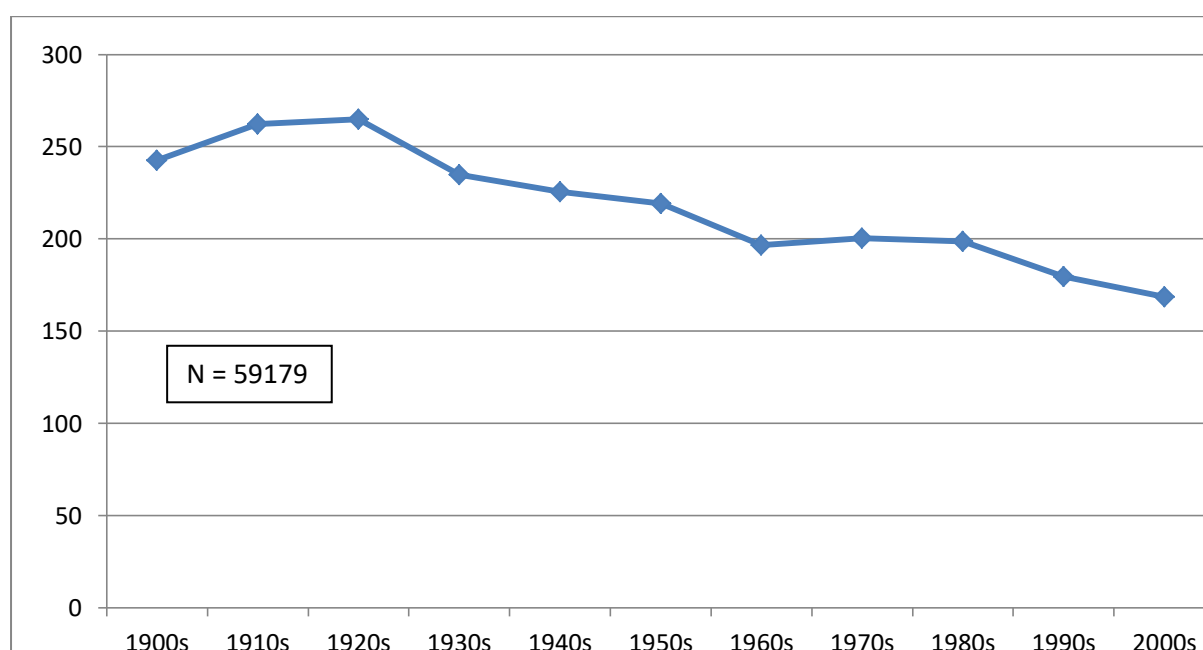


Figure 2: V + *that* entrenched by 1900 (24 verbs, pmw)

Another seven verbs did not meet my criterion for entrenchment in 1900 but became fairly frequent later in the century (*accept, concede, disclose, emphasize, establish, repeat, stress*); see Figure 3. They too showed a fall, if starting slightly later – after the 1960s or 1970s – apart from *accept*, which starts very low and rises rapidly after the 1960s. With this group, all examples were verified.

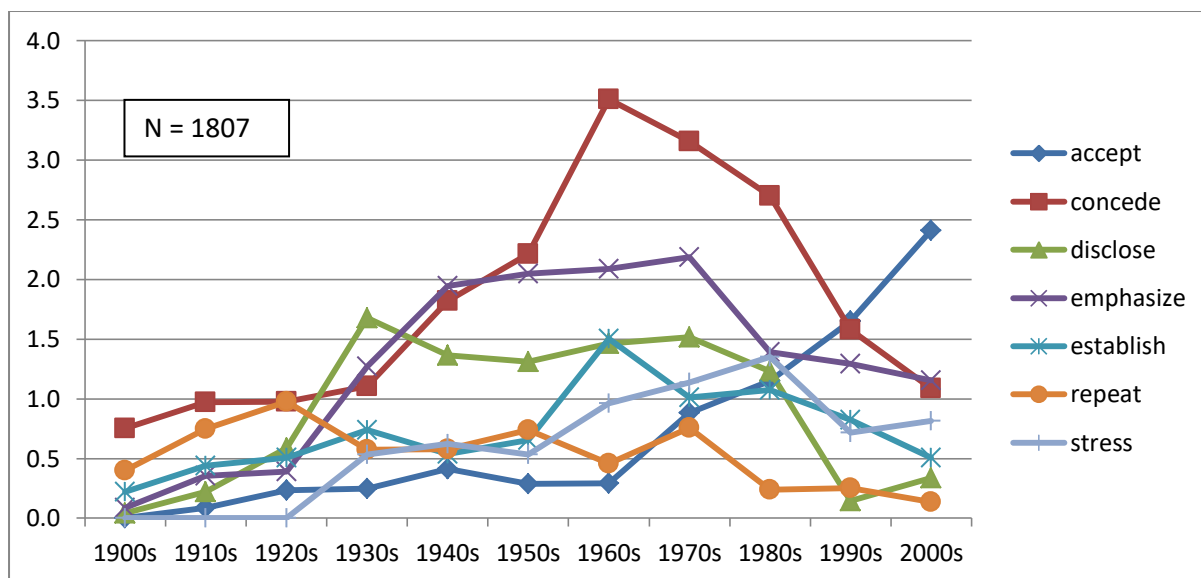


Figure 3: Medium-frequency verbs (7 verbs, pmw)

The remaining 137 verbs in the list had also produced few or no hits in the 1900s or 1910s, but genuine examples of V + *that* in later decades were not numerous. With these verbs the precision of the search was poor, and again all hits had to be checked and filtered. The totals per decade are shown in Figure 4.¹⁵ The most frequent of the 52 verbs charted are *appreciate* (72 examples 1900-2009), *lament* (52), *like* (52), *comprehend* (43), *advertize* (41) and *betray* (32).

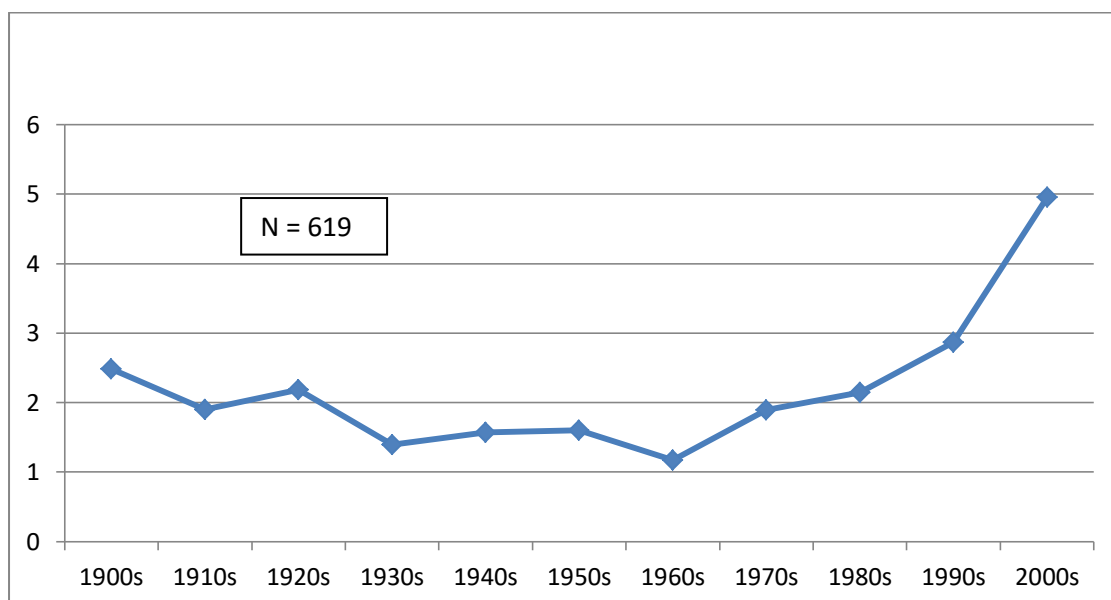


Figure 4: Low-frequency verbs (52 verbs, pmw)

¹⁵ The list is *advertize, appreciate, bear, betray, celebrate, change, comprehend, conceal, contemplate, contradict, corroborate, count, curse, deplore, develop, discount, disguise, dislike, disprove, dispute, disregard, echo, endorse, enjoy, entertain, face, formulate, hate, hide, ignore, include, lament, like, love, miss, offer, omit, overlook, ponder, posit, postulate, press, question, register, resent, sell, share, spot, symbolize, tout, underline, voice*.

There is a noticeable rise in the last four decades. To focus on this I select 12 verbs from the group (*appreciate, clarify, document, hate, highlight, ignore, like, love, miss, offer, posit, share*) plus *accept* (already charted in Figure 3 with the medium-frequency group) and plot them in Figure 5, showing raw numbers this time.

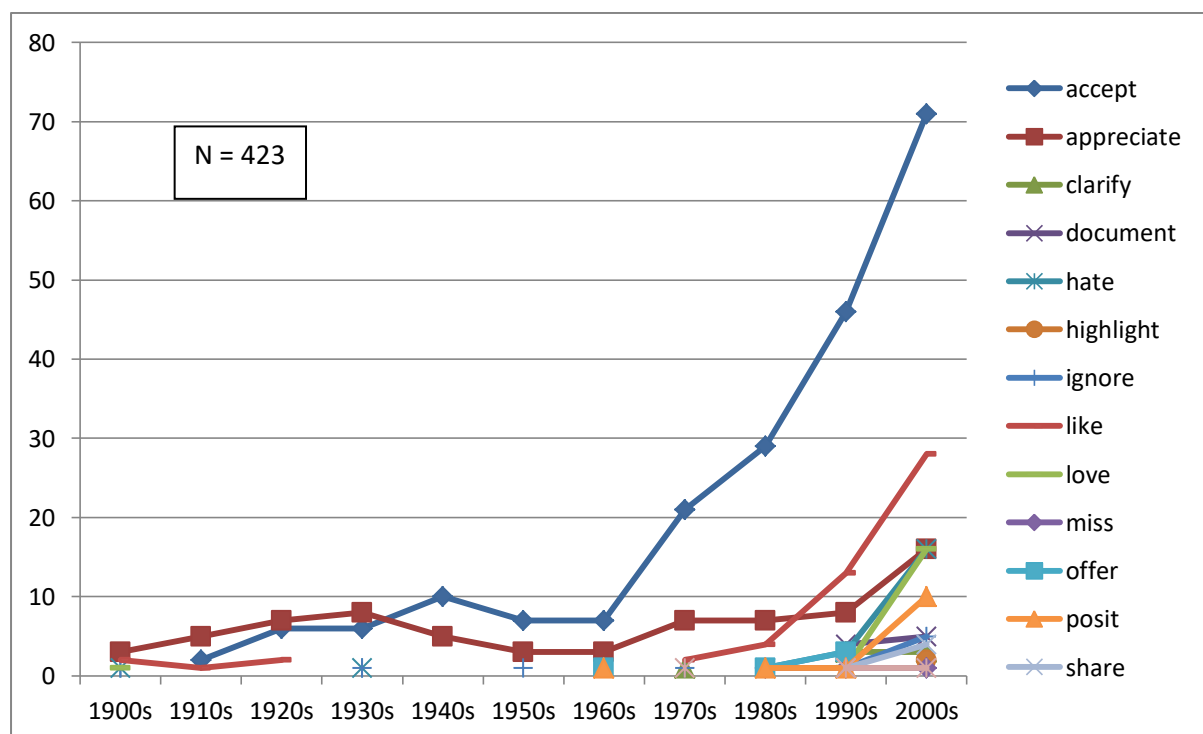


Figure 5: Recent uptick in V + *that* (13 verbs, raw frequencies)

Another ten verbs appear for the first time in V + *that* in either the 1990s or the 2000s, but with just one example (*disprove, disregard, entertain, spot, symbolize, tout, voice*) or two examples (*change, contradict, dislike*) each. Given more data, it might turn out that these verbs too are part of a recent uptick in V + *that*. I have left them out of Figure 5 to keep it readable. For the remaining verbs I have nothing to say about their individual trajectories .

In the light of these figures, what can we say about the history of V + *that* with our 98 verbs, all of which had appeared in shell + *that*? Note that the figures only represent instances found with one Python script, perhaps about two fifths of the instances in COHA. The general trend is a gentle decline, although the histories of quite a few verbs suggest a recent uptick in frequency, in some cases even an accretion to the construction. The verbs in Figure 5 (apart from *posit* and perhaps *document*) are everyday, colloquial words, mostly from a different register than those attested in my student essays.

There isn't enough information to mount a convincing constructional analysis that would explain specific rises in frequency or accretions to the class of verbs in *V + that*, when other verbs show no increase of usage or even (*consider, reflect*) a decline as we approach the present day. *V + that* is a long-established and compositional construction, a member of the more abstract transitive schema. In discourse terms, *V + that* and *shell + that* with a semantically neutral shell noun convey much the same, but *V + that* is more economical. It does not seem surprising that its distribution should widen.

I was to some extent able to test my own impressions for signs of the Recency Illusion, having casually assumed some examples to be new, whether classified as error or innovation. Some clearly were not new at all, for example *reflect + that*. Others may show an Anglo-American difference: an obvious candidate here is *like + that* (though that pattern had not appeared in my students' writing). Some patterns of *V + that* do seem to be recent, however, even in American English. For example, *criticise + that*, one of the usages that had struck me as new, occurs only once in COHA as far as I am aware (not, as it happens, in a pattern that my Python script picks out), and the date is indeed recent. Likewise *highlight + that* occurs only twice, again in the same decade.

(35)The shop teacher criticized that the two bookends were not perfectly symmetrical and gave him a B for the project (2002, COHA)

(36)The renowned Millennium Ecosystem Assessment [...]highlights that coasts are the most highly degraded ecosystems on our planet. (2005, COHA)

4 Whose grammar?

At some point in the investigation, explicit grammaticality judgements come into play, for example to exclude outright errors. However, such judgements are not always reliable. Although grammaticality judgements, even of sentences in isolation, can often achieve an uncontroversial consensus, outcomes may be influenced amongst other things by context, priming, dialect, idiolect, register, prescriptive ideas, and the Observer's Paradox. Much of this applies to linguists as well as everyone else. By definition, grammaticality must also figure in language change, since advanced and conservative speakers will differ in their (usually implicit) judgements.

How well-defined, then, is grammaticality? This is much discussed. To give three examples among many, Ross (1973: 389-91 [2004 reprint]) was an early commentator on variation of acceptability among speakers. More recently, Mindt (2002) argued that a corpus investigation of a grammatical rule will typically find up to 5% of exceptions due to language change, linguistic creativity, intended ungrammaticality, etc. (cited by Mukherjee 2005: 100-1); note that our problem is in part lexical, which Mindt excludes from his observation. As for Sampson & Babarczy (2014), they dispute the validity of the whole grammatical ~ ungrammatical distinction.

Within generative grammar there is a tradition of regarding graded judgements, whether from an individual speaker or the outcome of averaging the reports of many speakers, as a matter 'merely' of performance or E-Language and not part of the underlying grammar, which is assumed to be discrete and clear-cut (Schütze 1996, 2004). If gradience is acknowledged at all, then, it is what happens when subjective, extra-grammatical factors cloud the ability of speakers to report their own grammatical knowledge reliably.

The usage-based tradition, on the other hand, suggests that grammatical knowledge is inherently probabilistic (Gahl & Garnsey 2006, Gahl & Yu 2006, Bresnan 2007), which in some cases would disfavour or even preclude clear-cut judgements of grammaticality. I am sympathetic to this position, and indeed I regard speaker uncertainty as a useful diagnostic of usages that are marginal, problematic or undergoing change.

I return to the synchronic register contrast mentioned near the start of Section 1, and which is explained as follows by Biber et al. (1999: 650):

... the expression of stance is backgrounded in noun complement clauses. With verb complement clauses, the subject of the verb is often a human agent or experiencer, so that the stance reported by the verb can be attributed directly to that person. In contrast, the stance conveyed by a controlling head noun is not normally attributed to anyone, so that readers must infer that the noun reports the stance of the writer.

The opposite distributions of verb complement clauses (preferred in conversation) and noun complement clauses (preferred in academic writing) can thus be attributed to two factors. First, conversation has an overall preference for verbal rather than nominal structures, while academic prose shows the opposite pattern, preferring to integrate information in noun phrases. Second, the differing primary purposes of conversation and academic prose are at least as important here: conversational participants are interested in each others' [*sic*] personal feelings and attitudes [...] In contrast, academic writers are generally much more interested in the information being conveyed than personal attitudes [...]

In the case of what I have called 'unexpected V + *that*', my limited experience of asking students for a grammaticality judgement on what they had submitted is that they did not have strong intuitions either way. (Admittedly, given the situation, the Observer's Paradox may have contributed!) As unpractised writers, they may have been extending a verbal constructional pattern more typical of conversation to an academic context, but inserting lexical items into the construction that would be rare or non-existent in their everyday conversation. Do they really have control of the usage? To what extent is sporadic occurrence of 'new' V + *that* patterns a part of their grammar? It is almost as if they are not native speakers of the register of English they are learning to use. We can compare this sort of innovation with what would be regarded as errors in advanced L2 usage; there are affinities also with the development of new Englishes. See Hackert (2009) on the history and politics of the term 'native speaker', as contrasted with 'proficient speaker' (Mufwene 1998). This line of thinking is inconvenient for the idea of a speaker having one grammar, and of that grammar being well defined. It suggests too that research on L2 learning is relevant to this piece of diachronic research.

A more typical diachronic account would simply observe that some errors turn out with hindsight to have been innovations. Compare, say, the use of *criteria* as a singular instead of the 'correct' *criterion*. Evidently this represents a minor morphological change going on in the language. Similarly, the subject of this paper could also be seen as an unremarkable, language-internal extension of subcategorisation. I contend that the latter viewpoint is not mutually exclusive with the

previous discussion of registers and language learning.¹⁶ Our understanding of the V + *that* development can be enhanced by examining the social, technological and historical context and by considering register and medium.

Perhaps naively, I had not been sufficiently aware of the probable chronological difference between American and British developments before collecting data from COHA.¹⁷ It must be admitted that the COHA data are rather different from the British student writing I had first observed and offer little support to the hypothesis of ‘change by thesaurus’.

5 What else?

The wider historical context to the spread of V + *that* is *that*-clause complementation in general, which, as far as I know, is discussed mainly in relation to variation in two areas. The first is replacement by non-finite complements, specifically infinitives and *-ing* clauses (Rohdenburg 1995, Los 2005). Here the *that*-clause is the recessive variant, whereas we have been looking at a case where the *that*-clause *qua* verbal complement is the innovation. The difference, I suggest, is that non-finite clauses often involve a control or raising relationship between positions in higher and lower clause, whereas the verbs discussed in this paper are not open to subject raising, since they have agentive subjects, while the discourse function discourages a control relationship between higher and lower clause constituents.

The second area of variationist research is the choice between *that* and zero (Elsness 1984, Rissanen 1991, Thompson & Mulac 1991, Finegan & Biber 1995, López-Couso 1996). I have encountered very few examples of zero marking with the verbs I have been studying – what we might call V + $\emptyset$ as a variant of V + *that* – unsurprisingly, since my research method specifically targeted *that*. I suspect that the presence of V + $\emptyset$ could serve as a handy indirect indication of the prior entrenchment of V + *that* with the verb in question. This needs testing with a parsed corpus in order to seek out zero-marked clauses.

This paper has raised more questions than it can answer, and good answers would need more time and resources than the questions perhaps deserve. Even within the corpus linguistic approach there is a need to add sensitivity to genre and register, to develop more sophisticated search algorithms, to use larger historical datasets than COHA, and to systematically examine the Corpus of Current American English (COCA) and other contemporary corpora. Future work could perhaps take advantage of the better search facilities of BNCweb and compare BNC with BNC2014 (due for release later in 2017) for a recent snapshot of change, and in British English too. On genre it would be helpful to compare a corpus of student writing such as the Michigan Corpus of Upper-Level Student Papers (MICUSP) with standard corpora of (mainly) copy-edited material, in order to see whether the student work shows greater variability or innovation in V + *that*.

¹⁶ Register and domain even have a bearing on a ‘simple’ change like *criterion/criteria*, as that lexical item has arguably been moving from educated usage into such specialised discourses as computing and employment practice, and from there back to more general use.

¹⁷ When I presented on this topic at LMEC6 in Uppsala, Anne Curzan (American) observed that many of my ‘unexpected’ examples of V + *that* seemed quite normal to her, while Terry Walker (British) reported having been struck by the frequency of unexpected V + *that* while reading American young adult fiction.

The influence of technology on language is another area of research that could be brought to bear. Finally, and this is really a major desideratum, we need a psycholinguistic investigation of speakers' knowledge in order to put the discussion sketched in Section 4 on a sounder footing.

References

- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman grammar of spoken and written English*. Harlow: Pearson.
- Bresnan, Joan. 2007. Is syntactic knowledge probabilistic? Experiments with the English dative alternation. In Sam Featherston & Wolfgang Sternefeld (eds.), *Roots: Linguistics in search of its evidential base* (Studies in Generative Grammar 96), 75-96. Berlin and New York: Mouton de Gruyter.
- Davies, Mark. 2010-. COHA: The Corpus of Historical American English: 400 million words, 1810-2009. Available online at <http://corpus.byu.edu/coha/>. Brigham Young University.
- De Smet, Hendrik. 2012. The course of actualization. *Language* 88.3, 601-33.
- Denison, David. 2009. A new class of verbs taking *that*-clause complements. Paper presented at ICLCE3, Symposium on Current Change in the English Verb Phrase, London.
- Denison, David. 2011. ISLE highlights? Paper presented at ISLE2, Boston MA.
- Denison, David. 2017. *That*-clauses as complements of verbs or nouns. Paper presented at The Sixth International Conference on Late Modern English (LMEC 6): Internal and External Factors in Linguistic Stability and Language Change Uppsala.
- Elsness, J. 1984. *That* or zero? A look at the choice of object clause connective in a corpus of American English. *English Studies* 65, 519-33.
- Finegan, Edward & Douglas Biber. 1995. *That* and zero complementisers in Late Modern English: Exploring ARCHER from 1650-1990. In Bas Aarts & Charles Meyer (eds.), *The verb in contemporary English*, 241-57. Cambridge, etc: Cambridge University Press.
- Flowerdew, John & Richard W. Forest. 2015. *Signalling nouns in English: A corpus-based discourse approach*. Cambridge University Press.
- Gahl, Susanne & Susan Garnsey. 2006. Knowledge of grammar includes knowledge of syntactic probabilities. *Language* 82, 405-10.
- Gahl, Susanne & Alan C. L. Yu. 2006. Introduction to the special issue on exemplar-based models in linguistics. *The Linguistic Review* 23, 213-6.
- Hackert, Stephanie. 2009. A discourse-historical approach to the English native speaker. In Thomas Hoffmann & Lucia Siebers (eds.), *World Englishes – Problems, properties and prospects: Selected papers from the 13th IAWC conference* (Varieties of English Around the World G40), 385-406. Amsterdam and New York: John Benjamins.
- Hornby, Albert Sydney. 2015. *Oxford advanced learner's dictionary of current English*, 9th edn. Oxford University Press. <http://www.oxfordlearnersdictionaries.com/definition/english/>
- Huddleston, Rodney & Geoffrey K. Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
- Hunston, Susan & Gill Francis. 2000. *Pattern grammar: A corpus-driven approach to the lexical grammar of English* (Studies in Corpus Linguistics 4). Amsterdam: John Benjamins.
- López-Couso, María José. 1996. A look at *that/zero* variation in Restoration English. In Derek Britton (ed.), *English historical linguistics 1994: Papers from the 8th International Conference on English Historical Linguistics (8.ICEHL, Edinburgh, 19-23 September 1994)* (Current Issues in Linguistic Theory 135), 271-86. Amsterdam and Philadelphia: John Benjamins.
- Los, Bettelou. 2005. *The rise of the to-infinitive*. Oxford: Oxford University Press.
- Mindt, Dieter. 2002. What is a grammatical rule? In Leiv Egil Breivik & Angela Hasselgren (eds.), *From the COLT's mouth ... and others: Language corpora studies - in honour of Anna-Brita Stenström* (Language and Computers - Studies in Practical Linguistics 40), 197-212. Amsterdam and New York: Rodopi.

- Mufwene, Salikoko. 1998. Native speaker, proficient speaker and norms. In Rajendra Singh (ed.), *The native speaker: Multilingual perspectives*, 111-23. New Delhi: Sage.
- Mukherjee, Joybrato. 2005. *English ditransitive verbs: Aspects of theory, description and a usage-based model*. Amsterdam: Rodopi.
- Newmeyer, Frederick J. 2003. Theoretical implications of grammatical category-grammatical relation mismatches. In Elaine J. Francis & Laura A. Michaelis (eds.), *Mismatch: Form-function incongruity and the architecture of grammar* (CSLI Lecture Notes 163), 149-78. Stanford: University of Chicago Press for CSLI Publications.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London and New York: Longman.
- Rissanen, Matti. 1991. On the history of *that*/zero as object-clause links in English. In Karin Aijmer & Bengt Altenberg (eds.), *English corpus linguistics: Studies in honour of Jan Svartvik*, 272-89. London and New York: Longman.
- Rohdenburg, Günter. 1995. On the replacement of finite complement clauses by infinitives in English. *English Studies* 76, 367-88.
- Ross, John Robert. 1973. Nouniness. In Osamu Fujimura (ed.), *Three dimensions of linguistic theory*, 137-257. Tokyo: TEC for Tokyo Institute for Advanced Studies of Language. Repr. Aarts et al. (eds.) *Fuzzy Grammar* (2004), pp.351-422.
- Sampson, Geoffrey & Anna Babarczy. 2014. *Grammar without grammaticality: Growth and limits of grammatical precision* (Topics in English Linguistics (TiEL) 254). Berlin and Boston: De Gruyter Mouton.
- Schmid, Hans-Jörg. 2000. *English abstract nouns as conceptual shells: From corpus to cognition* (Topics in English linguistics 34). Berlin and New York: Mouton de Gruyter.
- Schütze, Carson T. 1996. *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago and London: Chicago University Press.
- Schütze, Carson T. 2004. The nature of graded judgments. In Bas Aarts, David Denison, Evelien Keizer & Gergana Popova (eds.), *Fuzzy grammar: A reader*, 431-46. Oxford: Oxford University Press. Repr. of Section 3.3 of Schütze (1996).
- Thompson, Sandra A. & Anthony Mulac. 1991. The discourse conditions for the use of the complementizer *that* in conversational English. *Journal of Pragmatics* 15, 237-91.
- Zwicky, Arnold. 2005. Just between Dr. Language and I. *Language Log*.
<http://itre.cis.upenn.edu/~myl/language-log/archives/002386.html> <Accessed 4 May 2006>.